

Proviso with Pseudowords: Binary + Scale

2020 Dec

Links:

1. Binary choice with null context: https://spellout.net/ibexexps/sherrychen/proviso2020_pilot4_bi_n/experiment.html
2. Binary choice with independent+dependent contexts: https://spellout.net/ibexexps/sherrychen/proviso2020_pilot4_bi_i-d/experiment.html
3. Likrt scale rating with null context: https://spellout.net/ibexexps/sherrychen/proviso2020_pilot4_likrt_n/experiment.html
4. Likrt scale rating with independent+dependent contexts: https://spellout.net/ibexexps/sherrychen/proviso2020_pilot4_likrt_i-d/experiment.html

Key Changes made from Pilot 3

- Context: simplified, somewhere between pilot 1 & pilots 2/3
- Training session: providing examples of contradictory, consistent, new but consistent
- Mixed design
 - Control + Null
 - Control + Independent + Dependent
- Two types of DVs:
 - Binary choice: Yes/No
 - Scale: 0-100

Changelog from Pilot 1 to Pilot 4

- Pilot 1: control+indep+dep*2, no context, 0-100 scale
 - Feels too much like a logical puzzle from the get-go
- Pilot 2: control+null+indep+dep*2, extensive context, 0-100 scale
 - Clear strategizing as evidenced by the increasing ratings for Ant across trials (which are expected to be the same across trial types)
- Pilot 3: control+null / control+indep / control+dep*2, extensive context, 0-100 scale
 - Totally between-subject, cannot tell how a participant might behave differently across context
 - Full-Dep condition not doing much for us, keep only one dep condition
 - Context may be too extensive → have obscured the task or made it too implicit?
- Pilot 4 (current): control+null / control+indep+dep, reduced context
 - Binary choice + 0-100 scale
 - Training session: provided clarity, reduced rejection rates

Participants

- Participants for each exp, recruited from Mturk
 - *Binary choice with null context: 28 recruited, 9 failing controls, **N=19***
 - *Binary choice with independent+dependent contexts: 25 recruited, 1 non-native + 4 failing controls, **N=20***
 - *Likert scale rating with null context: 24 recruited, 2 non-native + 3 failing controls, **N=19***
 - *Likert scale rating with independent+dependent contexts: 26 recruited, 6 failing controls, **N=20***
- Exclusions
 - With the training phase added, the exclusion rate is lower than before
 - Rejection rates are quite batch-dependent; need to be more careful with posting time

Binary Choice

1. *Binary choice with null context:*

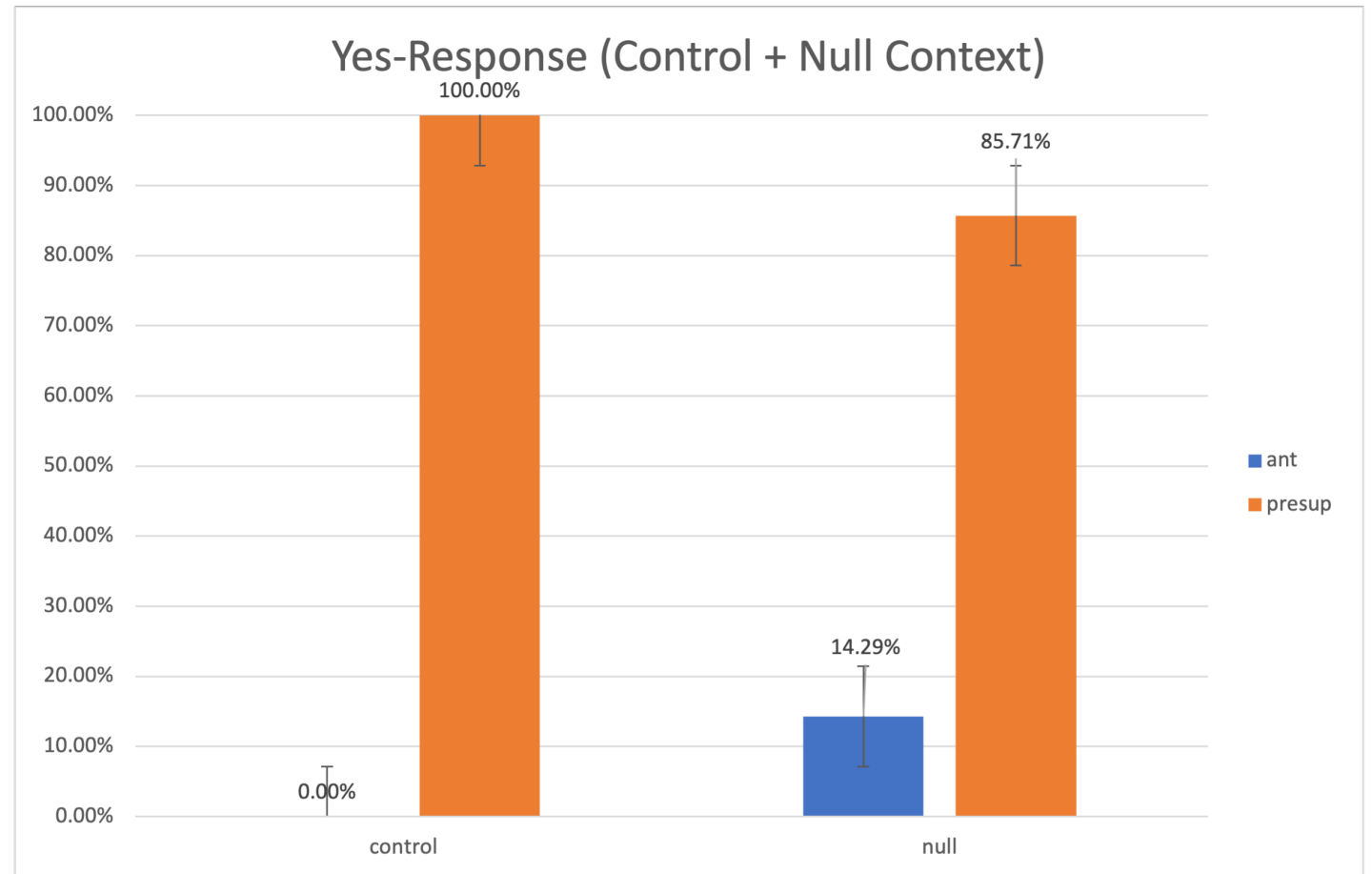
https://spellout.net/ibexexps/sherrychen/proviso2020_pilot4_bin/experiment.html

2. *Binary choice with independent+dependent contexts:*

https://spellout.net/ibexexps/sherrychen/proviso2020_pilot4_bi-i-d/experiment.html

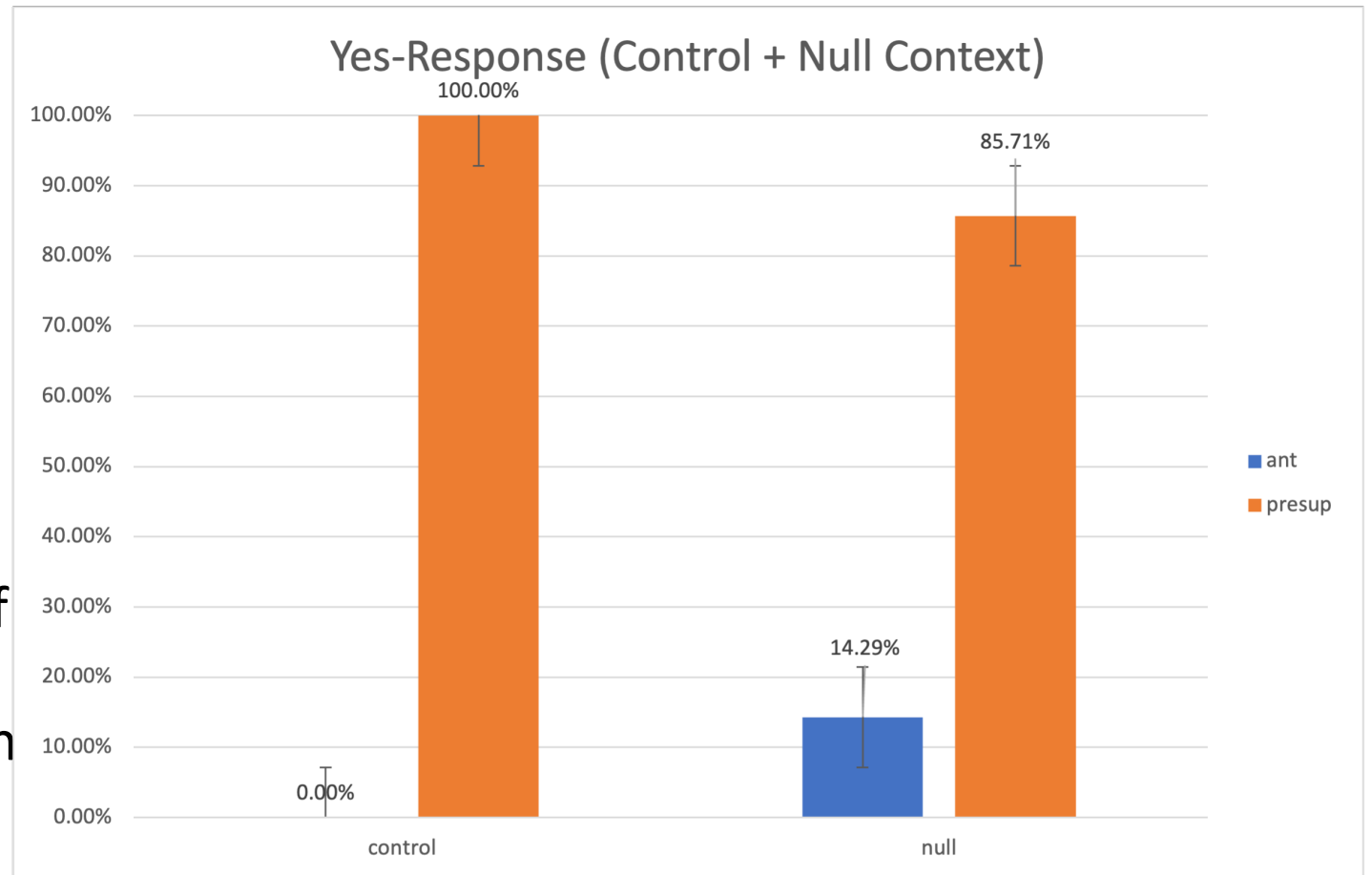
Binary choice: Control + Null

- Filtering criteria for **Control**:
 - Yes surprised to Presup “Eva doesn’t like to lammor”, since it contradicts the context
 - No surprised to Ant “Eva is a flepper”, since it’s given in the context



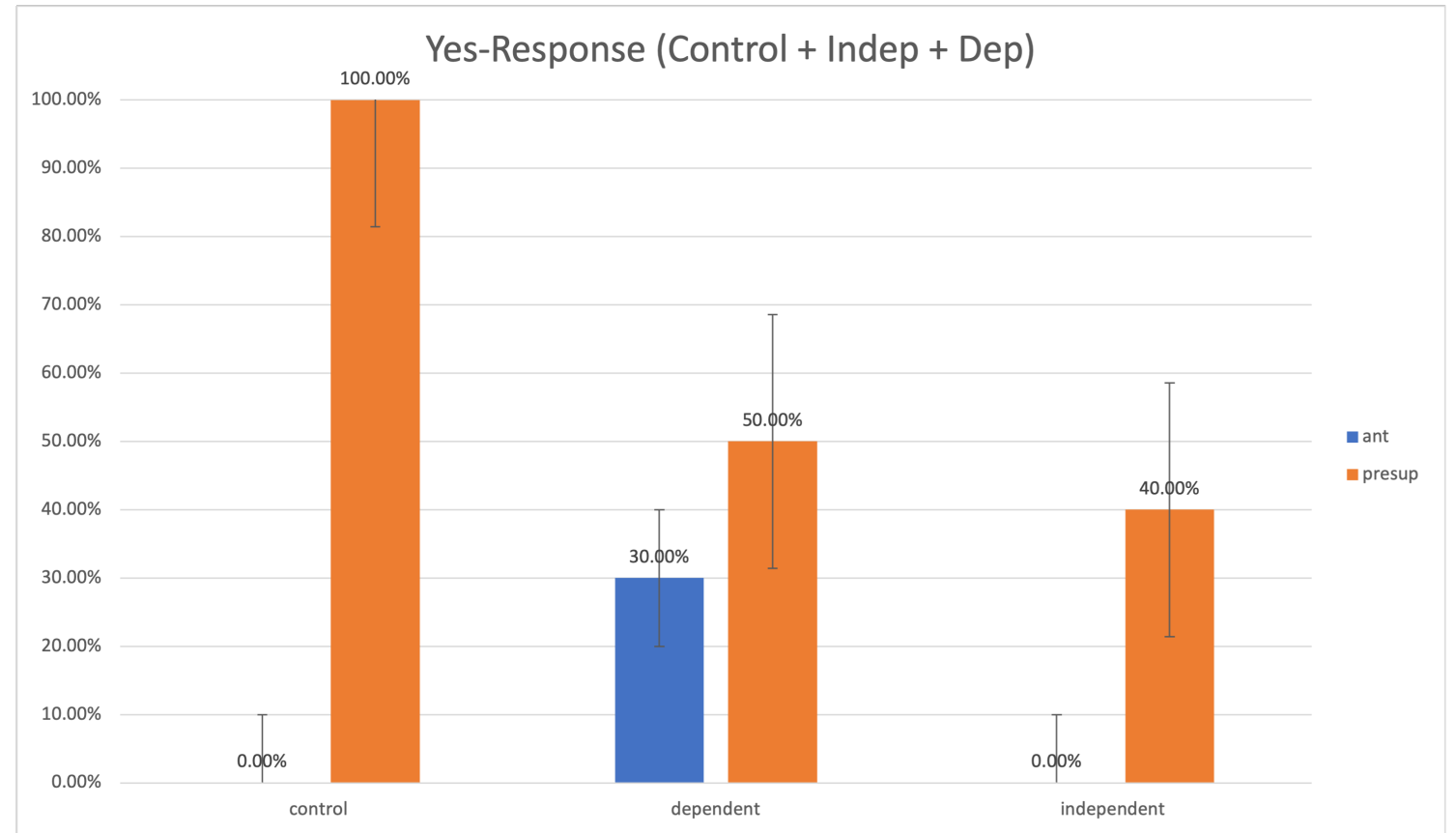
Binary choice: Control + Null

- Null Context:
 - Presup: strong preference to Yes response, close to Presup in Control; indicating commitment to the non-cond presup
 - Ant: low rates of Yes response, as expected (If X, then Y; nothing too surprising when we learn that X)



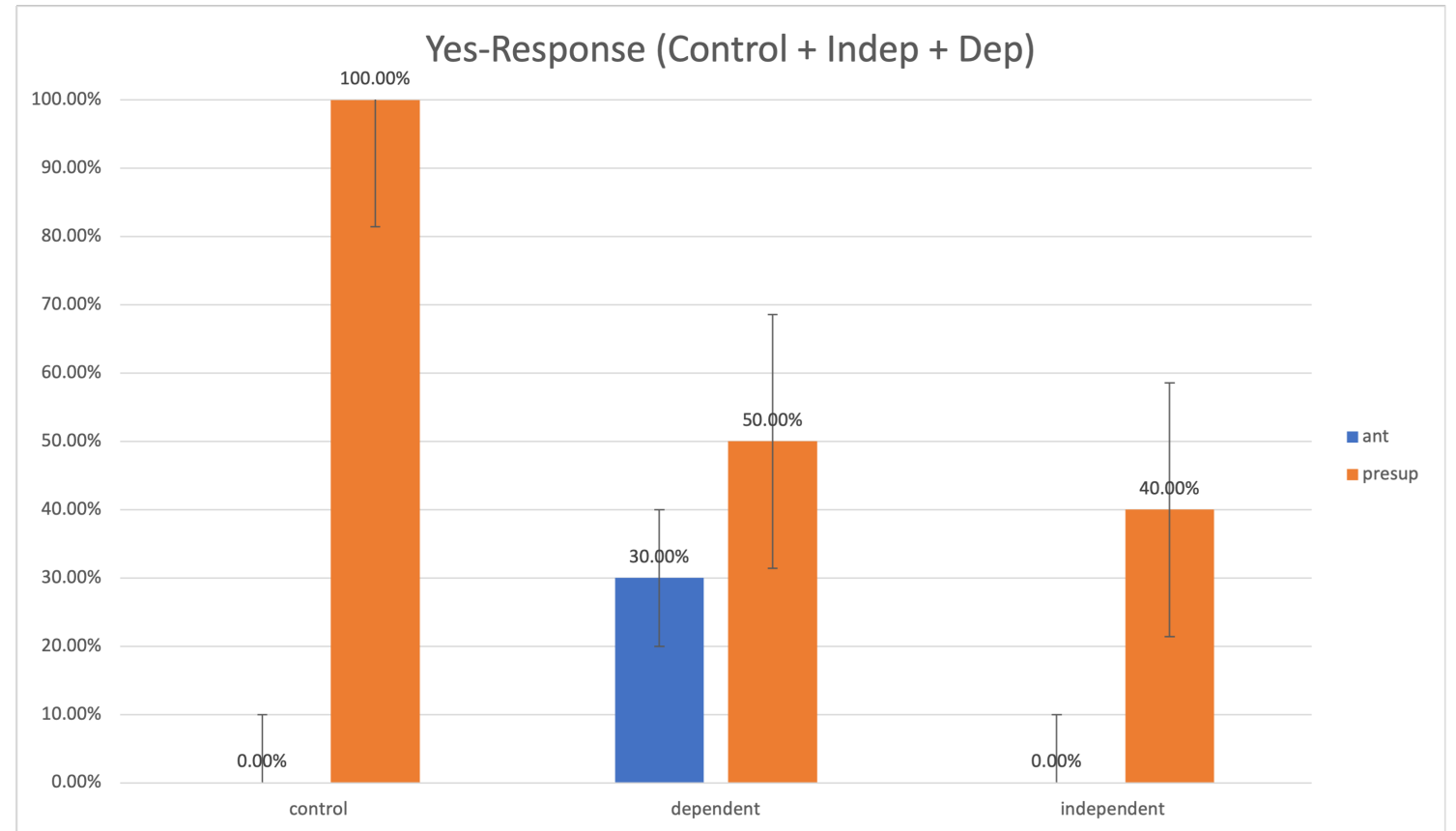
Binary choice: Control + Indep + Dep

- Filtering criteria for **Control**: same



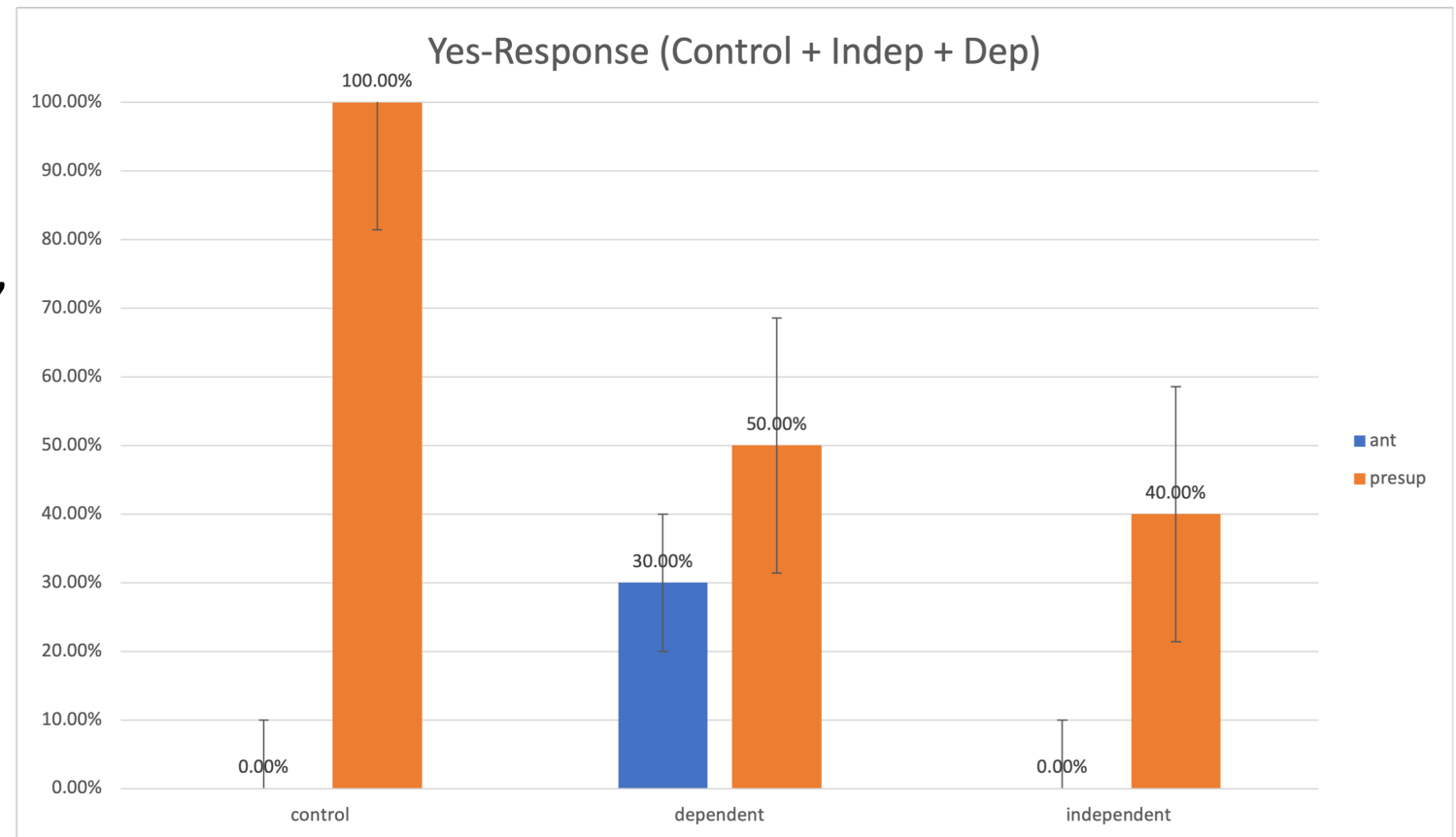
Binary choice: Control + Indep + Dep

- Independent:
 - 40% Yes to Presup;
Much lower than Null Context
 - 0% Yes to Ant, normal
- Dependent:
 - Presup: 50% Yes to Presup, no real difference with Indep (huge error bars!)
 - Ant: 30%??? Not normal!



Binary choice: Control + Indep + Dep

- Issue with Dependent Ant:
 - There are hints of this abnormality in previous pilots as well
 - Oddity: the instruction says **“Gogopos are known to usually have hooplers. No other aliens have hooplers.”**
 - After seeing Presup “Cece doesn’t have a hooper”, if ppl find that kinda surprising, then they do commit to Cece having a hooper. If so, then Cece has to be a gogopo. If so, they should **not** be surprised at Ant “Cece is a gogopo”!
 - On the other hand, if people only project the conditional presup, they also shouldn’t find Ant surprising.
- Weird order effect???



What is consistent with Pilots 2-3

- We see a difference between Null & Independent:
 - Pilots 2-3: ratings about 15 points higher for Null than Independent/Dep
 - Binary: 100% Yes for Null vs. 40% Yes for Indep
 - High Yes response rates for Null cannot be attributed to order effects (i.e. immediately following Control), because Indep also immediately follows Control

	Null	Independent	Dependent
Presup	85.71%	40%	50%
Ant	14.29%	0%	30%

0-100 Scale

3. Likrt scale rating with null context:

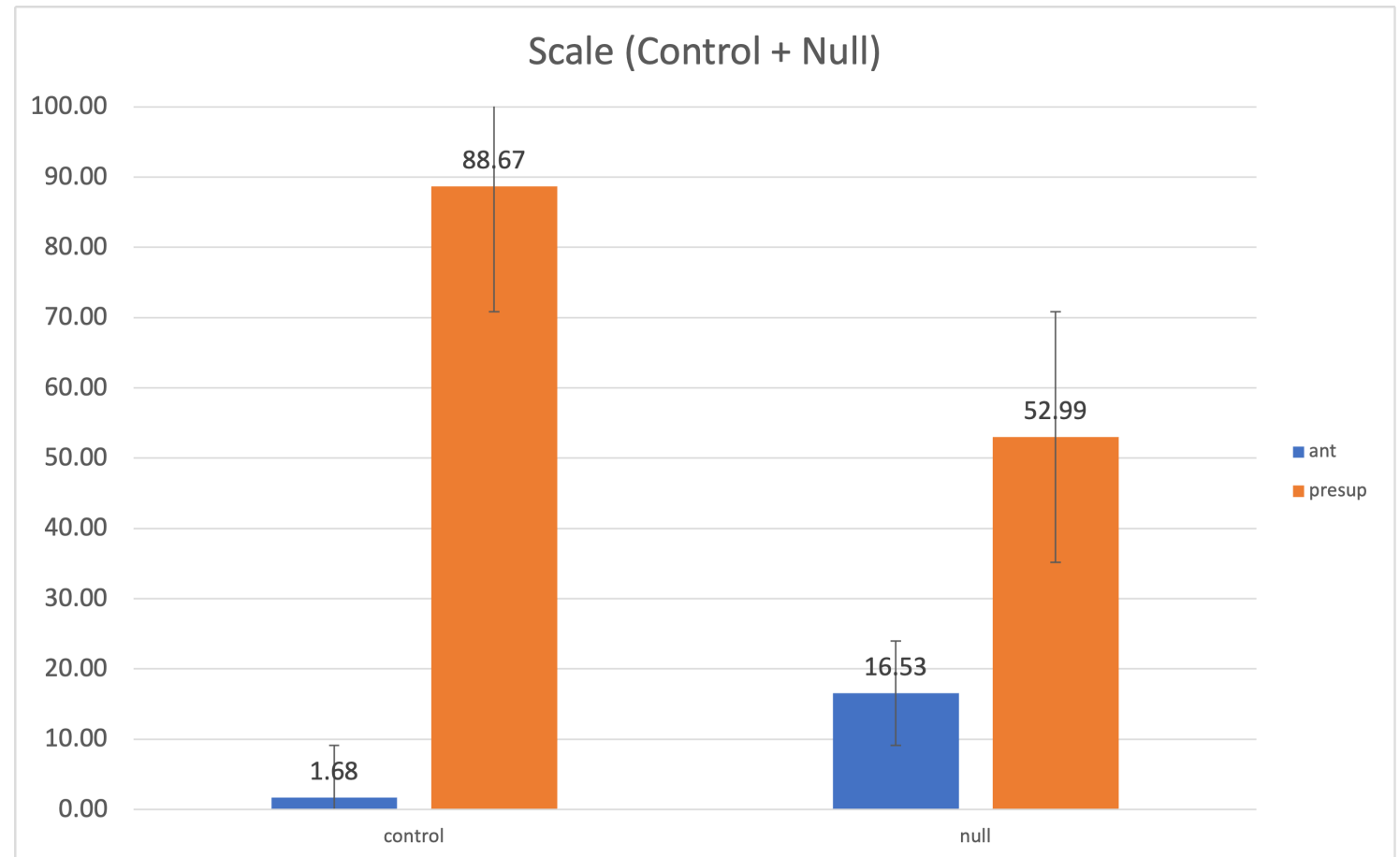
https://spellout.net/ibexexps/sherrychen/proviso2020_pilot4_likrt_n/experiment.html

4. Likrt scale rating with independent+dependent contexts:

https://spellout.net/ibexexps/sherrychen/proviso2020_pilot4_likrt_i-d/experiment.html

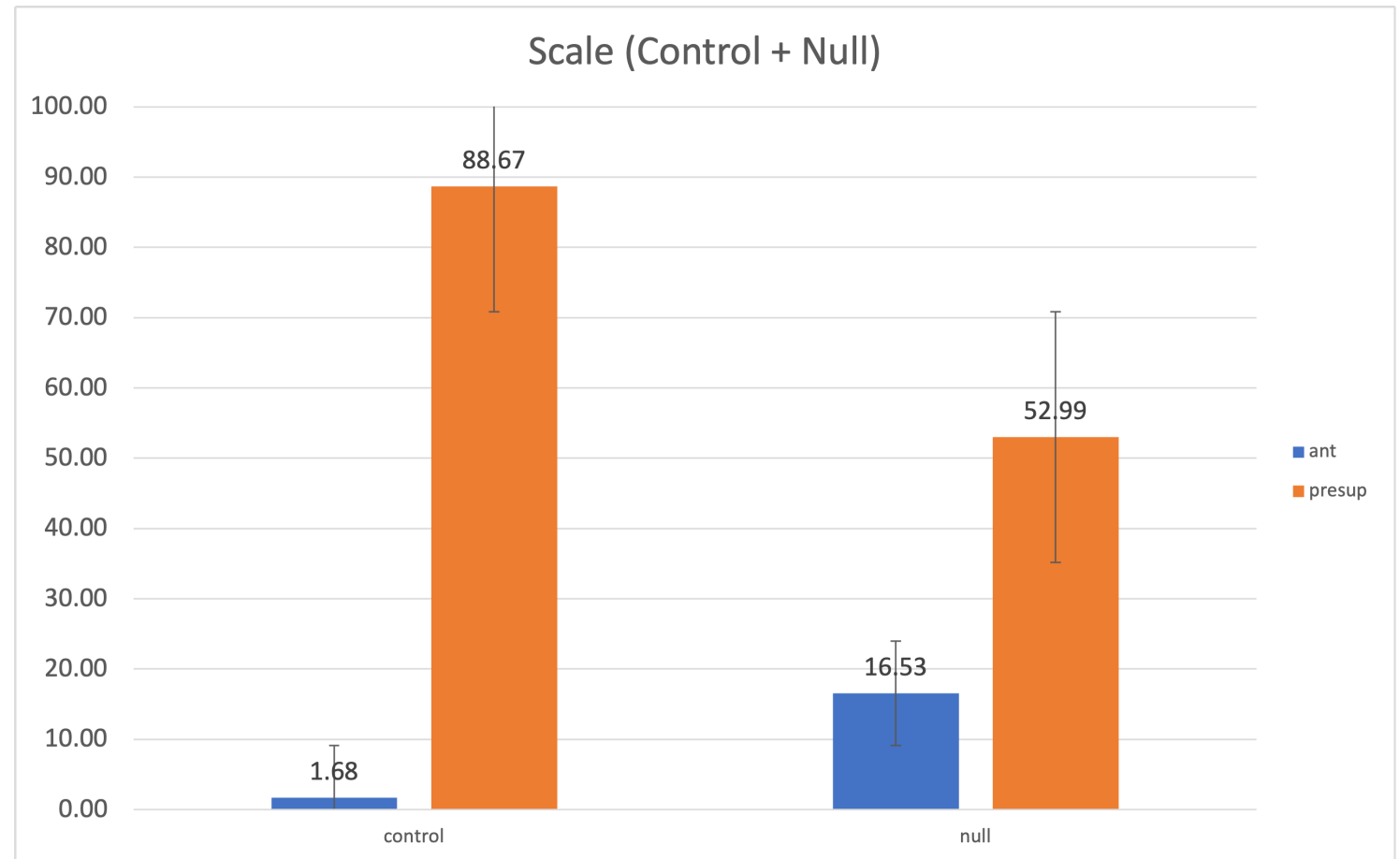
Ratings: Control + Null

- Filtering criteria for **Control**:
 - More surprised to Presup “Eva doesn’t like to lammor” than Ant “Eva is a flepper”, since the former is a contradiction and the latter given in the context



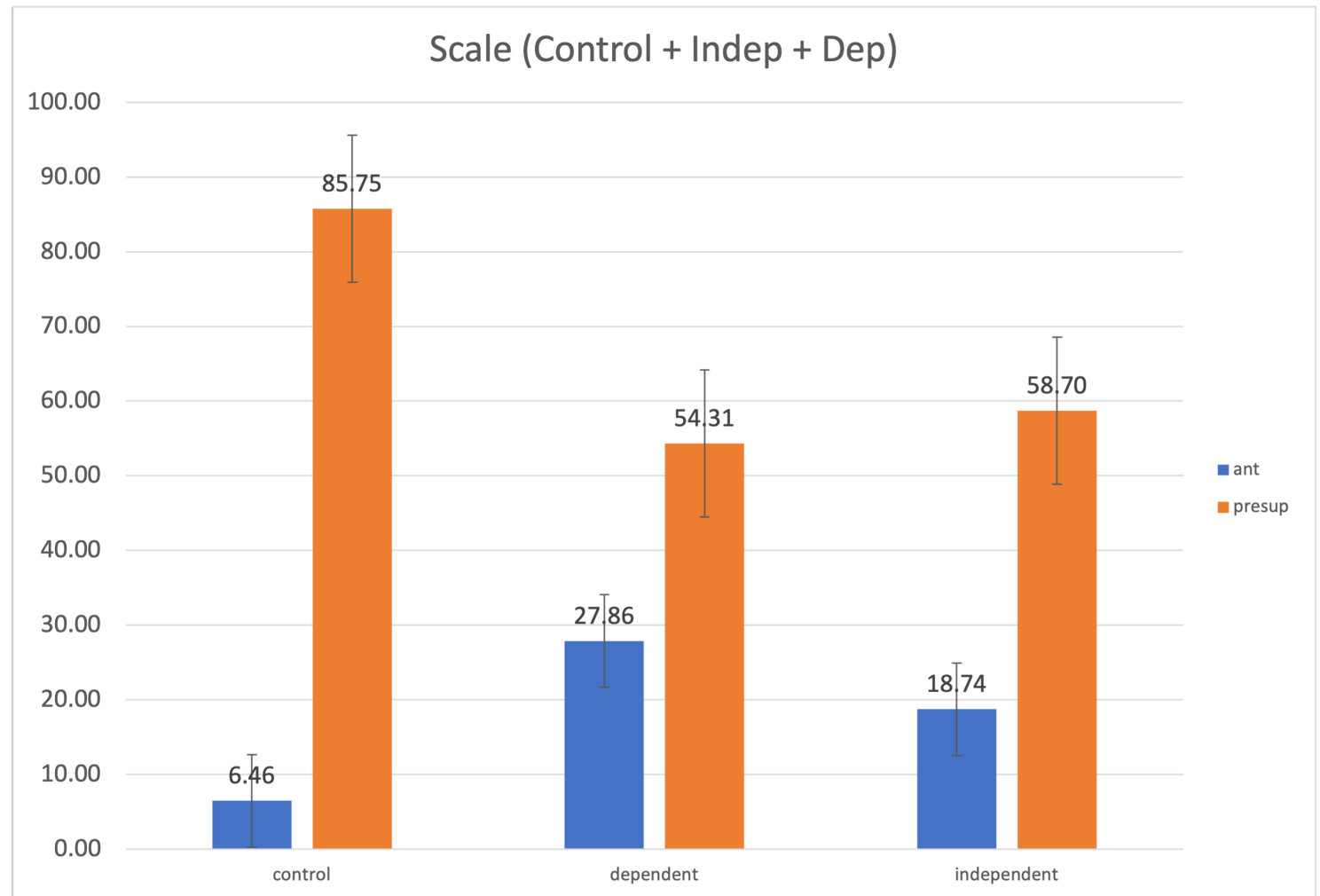
Ratings: Control + Null

- Null Context:
 - Rating for Presup
“John doesn’t have a dord”: 53, similar to previous pilots (56 in Pilot 2, 57 in Pilot 3)
 - Rating for Ant: low, as expected



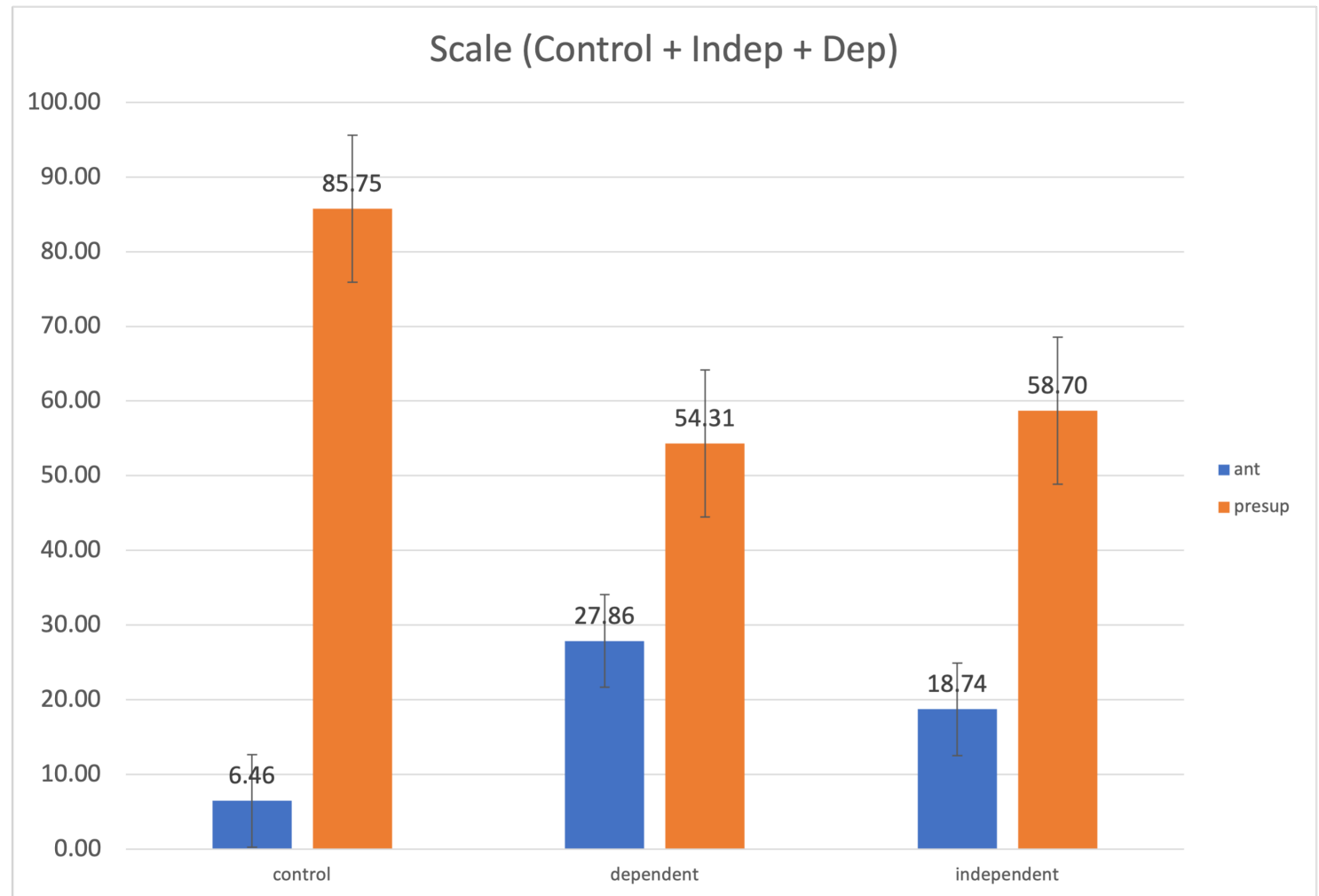
Ratings: Control + Indep + Dep

- Filtering criteria for **Control**: same



Ratings: Control + Indep + Dep

- Independent:
 - Presup: 58.7; this is considerably higher than ALL previous pilots (44, 46, 42)
- Dependent:
 - Presup: 54.3; no real diff from Indep
 - Ant: 28, also weirdly higher than the other Ant conditions.....



What is (in)consistent with Pilots 2-3

- Compared to in particular Pilot 3 (with full between-subject design)
 - Training in Pilot 4 brought down the ratings for Ant (clarity on what to do when you are just learning something new & non-contradictory, i.e. not very surprising)
 - All 3 critical conditions are rated quite similarly in Pilot 4

PILOT 3 Rating	Null	Independent	Dependent
Presup	57.95	42.51	50.14
Ant	35.32	33.67	36.42

PILOT 4 Rating	Null	Independent	Dependent
Presup	52.99	58.7	54.31
Ant	16.53	18.54	27.86

Where Binary & Scale results converged

- Lack of difference between Independent and Dependent
 - It does not seem like our independent manipulation is giving us much of a difference as we expected
- Weird behavior of Dependent-Ant
 - Seems to suggest that either (1) we did something wrong in that condition (no typo though), or (2) we really need to figure out the participants' rationale behind it that we are currently missing...

Yes-response	Null	Independent	Dependent
Presup	85.71%	40%	50%
Ant	14.29%	0%	30%

Ratings	Null	Independent	Dependent
Presup	52.99	58.7	54.31
Ant	16.53	18.54	27.86

Where Binary & Scale results differed

- Null Context Presup:
 - High Yes-response in Binary choice, close to the Control (involving a contradiction) and MUCH higher than Indep & Dep
 - But when rated on a 0-100 scale, not so different from Indep & Dep
 - What could have caused this difference between the two tasks?

Yes-response	Null	Independent	Dependent
Presup	85.71%	40%	50%
Ant	14.29%	0%	30%

Ratings	Null	Independent	Dependent
Presup	52.99	58.7	54.31
Ant	16.53	18.54	27.86

Recap of questions

- Why does Null Context received so many Yes-responses in Binary Choice, but is not rated very high on the Scale?
- The lack of independence effect in both tasks: Why is the independence manipulation not working?
- What's up with Dependent-Ant??? (What do responses to Ant tell us?)

Yes-response	Null	Independent	Dependent
Presup	85.71%	40%	50%
Ant	14.29%	0%	30%

Ratings	Null	Independent	Dependent
Presup	52.99	58.7	54.31
Ant	16.53	18.54	27.86